

Conditional Random Matrix Ensembles and the Stability of Dynamical Systems

Paul Kirk

Centre for Integrative Systems Biology and Bioinformatics, Imperial College London

E-mail: paul.kirk@email.com

Delphine M. Y. Rolando

Centre for Integrative Systems Biology and Bioinformatics, Imperial College London

E-mail: delphine.rolando10@imperial.ac.uk

Adam L. MacLean

Centre for Integrative Systems Biology and Bioinformatics, Imperial College London

E-mail: adam.maclea09@imperial.ac.uk

Michael P.H. Stumpf

Centre for Integrative Systems Biology and Bioinformatics, Imperial College London

E-mail: m.stumpf@imperial.ac.uk, ms.stumpf@imperial.ac.uk

Abstract. There has been a long-standing and at times fractious debate whether complex and large systems can be stable. In ecology, the so-called ‘diversity-stability debate’ [1] arose because mathematical analyses of ecosystem stability were either specific to a particular model (leading to results that were not general), or chosen for mathematical convenience, yielding results unlikely to be meaningful for any interesting realistic system. May’s work [2], and its subsequent elaborations, relied upon results from random matrix theory, particularly the circular law and its extensions, which only apply when the strengths of interactions between entities in the system are assumed to be independent and identically distributed (i.i.d.). Other studies have optimistically generalised from the analysis of very specific systems, in a way that does not hold up to closer scrutiny. We show here that this debate can be put to rest, once these two contrasting views have been reconciled — which is possible in the statistical framework developed here. Here we use a range of illustrative examples of dynamical systems to demonstrate that (i) stability probability cannot be summarily deduced from any single property of the system (e.g. its diversity), and (ii) our assessment of stability depends on adequately capturing the details of the systems analysed. Failing to condition on the structure of dynamical systems will skew our analysis and can, even for very small systems, result in an unnecessarily pessimistic diagnosis of their stability.

PACS numbers: 05.45.-a, 05.40.-a, 87.18.-h, 87.23.Cc

1. Introduction

While the notion of stability of the stationary solutions of dynamical systems has been particularly interesting (and divisive) in ecology, formal aspects of stability have also been studied extensively in other settings, notably engineering, (celestial) mechanics, the analysis of complex systems, and applied mathematics. For example, for linear time-invariant systems, the Routh-Hurwitz criterion sets out the conditions for global stability. More generally, the local stability of an equilibrium state of a non-linear ordinary differential equation (ODE) system can be assessed by inspecting the eigenvalues of the system's Jacobian matrix evaluated at the equilibrium [3]. If the real parts of the eigenvalues are all negative, then the equilibrium is (locally) stable. For any non-linear systems only local stability is implied by a negative leading eigenvalue. Given our interest in typically non-linear systems, we here consider the spectra of the Jacobians (or the sign of the largest eigenvalue, to be precise) directly, but keep in mind that the basin of stability may be finite, and potentially confined to a small region.

Following studies suggesting that complexity — or ecological diversity — was key to stability [4, 5], Gardner and Ashby, followed by May [6, 2, 7], considered how stability changes as complexity, defined in terms of the number of state variables (i.e., in the cases they consider, the number of species) and the probability of an interaction between two variables, increases. In order to generalise this analysis, instead of focusing on specific examples, May considered ensembles of Jacobians defined in terms of matrix probability distributions. For suitably defined random matrix ensembles (RME) [8, 9, 10] he showed that sufficiently large or complex systems have a probability of stability close to zero. Subsequent studies considered different RMEs designed to reflect a variety of features found in real systems, and have drawn different or more nuanced conclusions regarding stability [11, 12, 13]. Other authors have pointed out the lack of realism of this approach [14] and estimated stability probabilities for specific ODE systems, either through experiments [15, 16] or by sampling values for the system's parameters and — for each sample — identifying the equilibrium points and determining their stability [17, 22, 23, 24, 25, 26, 27, 28]. By repeatedly sampling in this way, Monte Carlo estimates of stability probabilities may be obtained. The advantage of such Monte Carlo approaches is that it is possible to condition on properties such as the *feasibility* of equilibrium points (i.e. whether or not they are physically meaningful), which can again yield different conclusions regarding stability [17]. These approaches also define RMEs, but do so implicitly and with reference to specific ODE models. Given the variety of conclusions that have been drawn by different authors, the choice of RME is clearly crucial in determining the stability probability.

Random matrices have also, of course, a distinguished track-record in different branches of physics. Following Wigner's earliest work on calculating fluctuations in the eigenvalue spectra of Hamiltonians describing atomic nuclei [19], they have found use in the analysis of a whole range of fluctuations in different application areas: solid state physics, chemical reactions and transition state theory, and quantum chaos are only

some of the areas where they, in particular in the guise of the *Gaussian Orthogonal Ensemble* (GOE) and its generalizations, have come to use. But RMEs have also been found useful in pure mathematics [20, 21], and, for example, the spectral properties of the *Gaussian Unitary Ensemble* capture the statistical properties of the zeros of Riemann's zeta function; more recently they have also been employed in cryptography.

In all these applications RMEs are used to describe fluctuations which are believed to be separable from the secular dynamics of the underlying system. Here our use is subtly different. Instead of considering RMEs as general descriptors of some system — this has also been the strategy of May and, perhaps to a lesser extent, his followers — we are trying to condition the RME on the properties of real systems that determine whether or not the stationary states are stable or not. This then allows us to calculate a probability for a system to become unstable upon a small but finite perturbation. So, rather than making general statements about stability, our RMEs, which we refer to as *conditional RMEs*, are explicitly geared towards the use in specific contexts. While the success of traditional RMEs in capturing universal dynamics is based on assuming symmetries and homogeneity in the matrix entries, the stability analysis of specific real-world systems requires our conditional RMEs to exhibit the same heterogeneities that characterize real-world (i.e. problem-derived) Jacobian matrices. We will show below that this is necessary to understand when and why a large dynamical system can be stable, but that this *fully conditioned RME* should not be used to draw general conclusions as any rule from this system-specific approach can only highlight the behaviour of that system or systems with similar dynamics.

2. Stability and Random Matrix Ensembles

For any particular parametric ODE model, the Jacobian matrix will usually exhibit structure and dependency between its entries, and will typically be a function of the model parameters and the state variables. The present work addresses the question of how assessments of stability change when the structure and dependency present in the Jacobian is properly taken into account.

For example, for the Lorenz system of ODEs (see *Appendix B* for details), the Jacobian is given by,

$$J(\sigma, r, b, x, y, z) = \begin{pmatrix} -\sigma & \sigma & 0 \\ r - z & -1 & -x \\ y & x & -b \end{pmatrix}.$$

As a consequence of this structure and dependency, and regardless of how we choose the parameters of the system, only a particular family of $n \times n$ real matrices will be obtainable as Jacobians of the system. For example, no matter what values we take for the parameters of the Lorenz system, the (1,3)-entry of the Jacobian matrix will always be zero, and the (1,1)-entry will always be equal to the negative of the (1,2)-entry. It follows that if we were interested in assessing the stability probability of

one of the Lorenz system's equilibrium points, it would be inappropriate to calculate $P(\text{stable}|h)$ using, for example, a matrix probability density function, h , that associates non-zero density with matrices for which the (1,3)-entry is non-zero or $(1,1) \neq -(1,2)$. Nevertheless, many previous analyses have failed to account for the structure and dependency present in realistic Jacobian matrices (i.e. ones derived from models of real systems), instead restricting attention to matrix probability density functions that yield analytically tractable results and assuming that the results so obtained were general.

2.1. An Illustration

To further illustrate the implications of neglecting Jacobian structure, we consider a number of examples from a family of ODEs whose Jacobians have the form $\begin{pmatrix} -1 & a \\ b & -1 \end{pmatrix}$, with $a, b \in \mathbb{R}$. In this case, the space of all matrices can be straightforwardly represented as a 2-dimensional Cartesian coordinate plane, in which the abscissa describes the value taken by a and the ordinate the value taken by b (as in Fig. 1).

More precisely, we consider systems of the form,

$$\begin{aligned} \frac{dx}{dt} &= -x + g_1(y, \theta) \\ \frac{dy}{dt} &= -y + g_2(x, \theta) \end{aligned} \tag{1}$$

whose Jacobians are given by,

$$\begin{pmatrix} -1 & \frac{\partial}{\partial y} g_1(y, \theta) \\ \frac{\partial}{\partial x} g_2(x, \theta) & -1 \end{pmatrix},$$

where θ is the vector of model parameters, and x and y are the state variables.

Example 1: We start by considering the following (simple, linear) choices for g_1 and g_2 :

$$\begin{aligned} g_1(y, \theta) &= \theta_1 y \\ g_2(x, \theta) &= \theta_2 x, \end{aligned}$$

with θ_1 and θ_2 both non-zero. In this case, the Jacobian for the system is

$$\begin{pmatrix} -1 & \theta_1 \\ \theta_2 & -1 \end{pmatrix},$$

which is a function only of θ_1 and θ_2 (and not of x or y).

The equilibrium points are given by solving the simultaneous equations:

$$\begin{aligned} \frac{dx}{dt} = 0 &\Rightarrow -x + \theta_1 y = 0 \\ \frac{dy}{dt} = 0 &\Rightarrow -y + \theta_2 x = 0. \end{aligned}$$

It is straightforward to show that the only solution that holds for all values of θ_1, θ_2 is $[x, y] = [0, 0]$. For this system, the form of the Jacobian means that we may obtain *any*

matrix of the form $\begin{pmatrix} -1 & a \\ b & -1 \end{pmatrix}$, provided that each of the parameters θ_1 and θ_2 can take any value in \mathbb{R} . We do note, however, that in practice this model is likely to be of only limited interest, since it describes a system in which both of the interacting variables (e.g. species) will eventually become extinct.

Example 2: We next consider a nonlinear example:

$$\begin{aligned} g_1(y, \theta) &= \theta_1 y^2 \\ g_2(x, \theta) &= \theta_2 x, \end{aligned}$$

with θ_1 and θ_2 both non-zero. In this case, the Jacobian for the system is

$$\begin{pmatrix} -1 & 2\theta_1 y \\ \theta_2 & -1 \end{pmatrix},$$

which is a function not only of θ_1 and θ_2 , but also y .

It is straightforward to show that the only equilibrium points (EPs) that exist for all permitted values of θ_1 and θ_2 are: (i) EP1: $[x, y] = [0, 0]$; and (ii) EP2: $[x, y] = \left[\frac{1}{\theta_1 \theta_2^2}, \frac{1}{\theta_1 \theta_2} \right]$.

The Jacobian evaluated at EP1 is $\begin{pmatrix} -1 & 0 \\ \theta_2 & -1 \end{pmatrix}$. Thus the region of the (a, b) Cartesian coordinate plane representing the possible Jacobians associated with EP1 is simply the line $a = 0$. Similarly, the Jacobian evaluated at EP2 is $\begin{pmatrix} -1 & 2/\theta_2 \\ \theta_2 & -1 \end{pmatrix}$, and hence the region representing the possible Jacobians associated with EP2 is the line $b = 2/a$.

Example 3: We consider a further nonlinear example:

$$\begin{aligned} g_1(y, \theta) &= \theta_1 y^3 \\ g_2(x, \theta) &= \theta_2 x, \end{aligned}$$

with θ_1 and θ_2 both non-zero. In this case, the Jacobian for the system is

$$\begin{pmatrix} -1 & 3\theta_1 y^2 \\ \theta_2 & -1 \end{pmatrix},$$

which is again a function of θ_1, θ_2 , and y .

In this case, it is again straightforward to show that the only equilibrium points (EPs) that exist for all permitted values of θ_1 and θ_2 are: (i) EP1: $[x, y] = [0, 0]$; and (ii) EPs 2 and 3: $[x, y] = \pm \left[\frac{\theta_1}{\sqrt{\theta_1^3 \theta_2^3}}, \frac{1}{\sqrt{\theta_1 \theta_2}} \right]$.

The Jacobian evaluated at EP1 is again $\begin{pmatrix} -1 & 0 \\ \theta_2 & -1 \end{pmatrix}$. Thus the region of the (a, b) Cartesian coordinate plane representing the possible Jacobians associated with EP1 is again the line $a = 0$. The Jacobian evaluated at EP 2 or 3 is $\begin{pmatrix} -1 & 3/\theta_2 \\ \theta_2 & -1 \end{pmatrix}$, and hence the region representing the possible Jacobians associated with these EPs is the line $b = 3/a$.

Assessing stability for these examples: For any matrix of the form $J = \begin{pmatrix} -1 & a \\ b & -1 \end{pmatrix}$, the characteristic equation $|J - \lambda I_2| = 0$ may be expanded as $(-1 - \lambda)(-1 - \lambda) - ab = 0$, i.e. $(\lambda + 1)^2 - ab = 0$. The eigenvalues of J are the solutions of this equation, and are given by $\lambda_{1,2} = -1 \pm \sqrt{ab}$. J is in the stable region, Λ_S^2 , provided the real parts of $\lambda_{1,2}$ are both negative. First, we note that if ab is negative (i.e. if $\text{sgn}(a) = -\text{sgn}(b)$) then the real part of both eigenvalues is -1 and hence J is in the stable region. If ab is positive, then $\lambda_2 = -1 - \sqrt{ab} < -1$, so this eigenvalue is certainly negative, and it remains only to consider the sign of the other eigenvalue, $\lambda_1 = -1 + \sqrt{ab}$. This eigenvalue is negative if and only if $ab < 1$. We may thus completely determine the stable region for matrices of the form $\begin{pmatrix} -1 & a \\ b & -1 \end{pmatrix}$, as illustrated in Fig. 1. We also show the regions representing the Jacobians evaluated at the equilibrium points for the systems considered in Examples 1 – 3. Wherever these regions intersect the stable region, the corresponding equilibrium point(s) will be stable. The *probability* of a particular system being stable around a given equilibrium point, \mathbf{x}_0 , is therefore equivalent to the probability of the relevant Jacobian evaluated at \mathbf{x}_0 falling within one of these intersections. We consider how this probability should be defined in the next section. However, it is clear that if we ignore the existence of these regions when defining the matrix probability density function denoted h in Equation (2), and instead choose h in an arbitrary manner for the sake of analytical tractability (as illustrated by the contour lines in Fig. 1), then the resulting value we obtain for the “stability probability” will be similarly arbitrary, and hence have little meaning or validity for any specific model.

2.2. Formal description

For any system of n ordinary differential equations (ODEs), the Jacobian matrix of the system evaluated at a particular equilibrium point \mathbf{x}_0 will be an element, J , of the set of $n \times n$ real matrices $M_n(\mathbb{R})$. The equilibrium point \mathbf{x}_0 is locally stable if all of the eigenvalues of J have negative real part. An equivalent criterion is that the real part of the leading eigenvalue (i.e. the one having maximal real part) is negative.

We first consider the set of all $n \times n$ real matrices, $M_n(\mathbb{R})$. The eigenvalues of any matrix $J \in M_n(\mathbb{R})$ are the solutions of the characteristic equation,

$$|J - \lambda I_n| = 0,$$

where I_n denotes the $n \times n$ identity matrix [3]. We define $\Lambda_S^n \subset M_n(\mathbb{R})$ to be the set of $n \times n$ matrices having all negative eigenvalues, and refer to this as the *stable region* of $M_n(\mathbb{R})$.

The choice of a particular RME specifies a probability density function, h , on $M_n(\mathbb{R})$. The stability probability associated with h is then

$$P(\text{stable}|h) = \int_{J \in \Lambda_S^n} h(J) dJ, \quad (2)$$

i.e. it is the total probability mass that falls within the stable region.

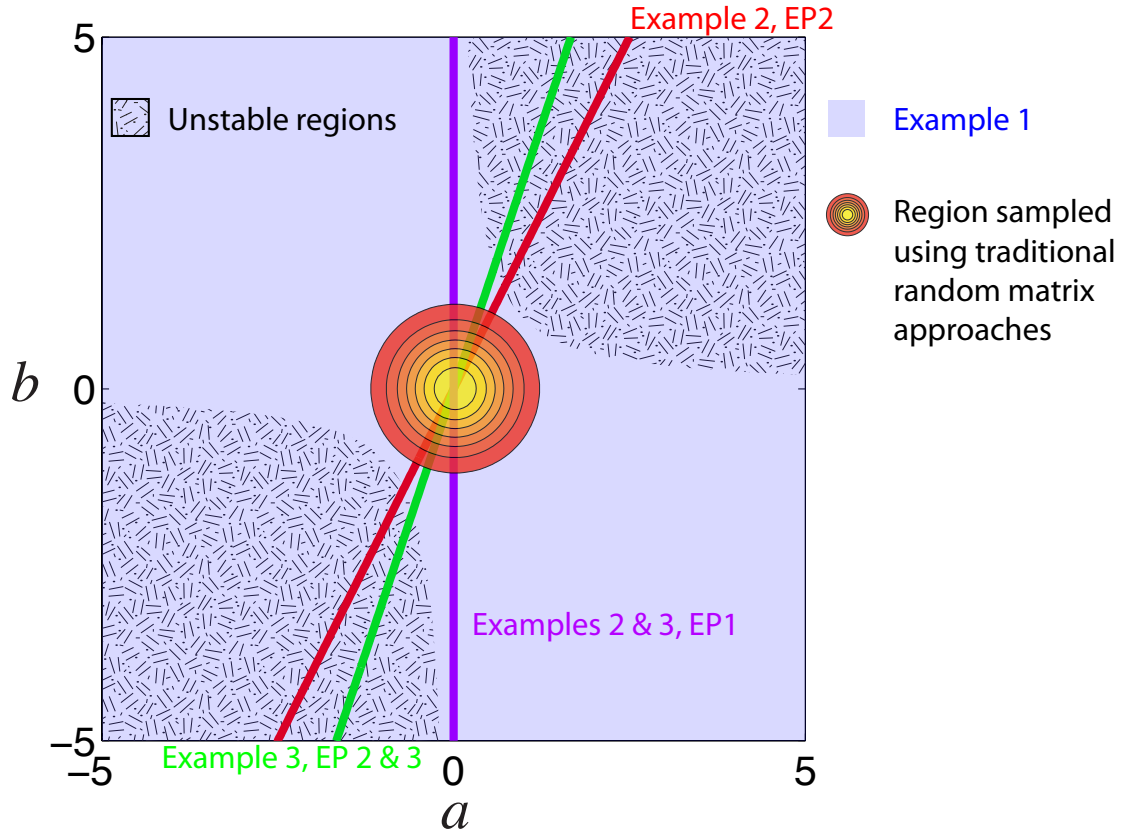


Figure 1. Stability of example equilibrium points (EPs) of ODEs. We consider different values for a and b and show as hatched areas (labelled “unstable regions”) the regions of the plane for which the resulting matrix has an eigenvalue with non-negative real part. The non-hatched area corresponds to the stable region for matrices of the form $\begin{pmatrix} -1 & a \\ b & -1 \end{pmatrix}$. We illustrate regions of the plane which correspond to the Jacobians that may be obtained for the various ODE systems and equilibrium points considered in Examples 1–3 (blue shaded area, and red, green, and purple lines, as indicated). We also represent using contours the random matrix distribution that has traditionally been considered in the literature when assessing stability probabilities.

Crucially, the stability probability is determined by two factors: Λ_S^n and h . Λ_S^n is not random: for a given n it is a well-defined region of $M_n(\mathbb{R})$. For systems of practical interest, however, this area cannot be determined analytically, and needs instead to be evaluated computationally, using e.g. Monte Carlo techniques (which are outlined below in 2.3). The results of stability analyses will therefore be completely determined by the choice of h , and how it distributes probability mass over the stable and unstable regions (see Fig. 2 and Section 2). If h is defined through the specification of an ODE model and a distribution for its parameters, then only matrices that can occur as Jacobians for that particular system will have non-zero density. If h is this time defined *without* reference to a real system, then, similarly, only some of the matrices in $M_n(\mathbb{R})$ will be associated with non-zero density, however, for any given real ODE system, these matrices might not be obtainable as Jacobians, and – conversely – not all matrices

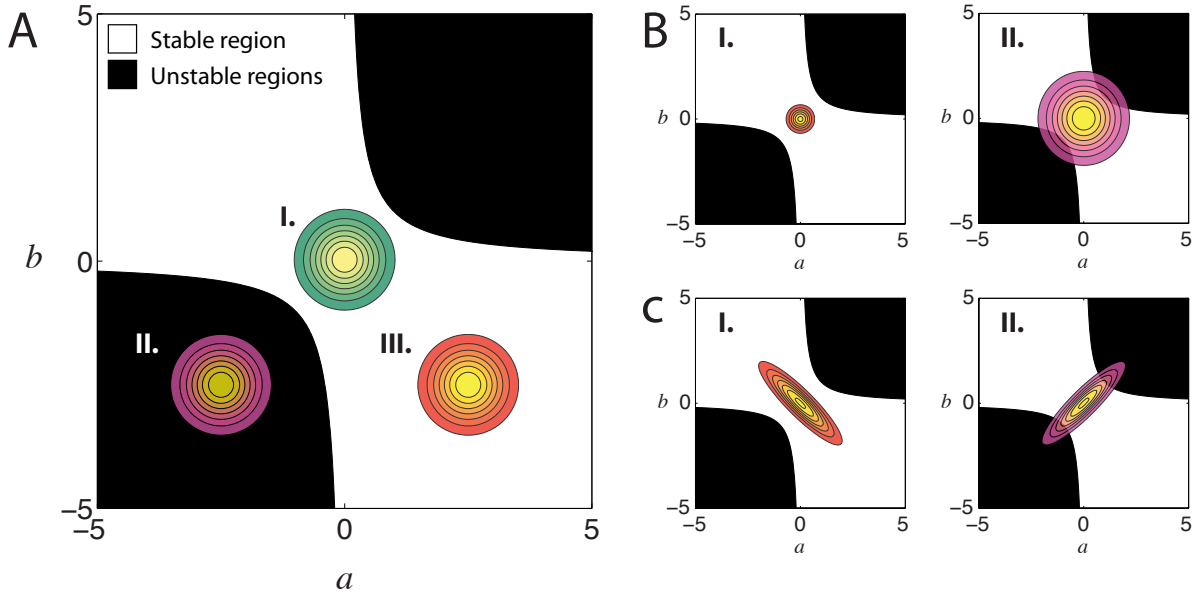


Figure 2. The means, variances and covariances of entries of random Jacobian matrices all have an impact upon stability probability. To illustrate, we consider 2×2 random matrices with off-diagonal terms $[a, b]^\top \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and -1 on the diagonal. It is straightforward to show (see *Appendix*) that such matrices are stable if $ab < 1$ and unstable if $ab > 1$. We take various choices for $\boldsymbol{\mu}$ and Σ , and illustrate the resulting bivariate normal distributions using coloured contours. A. The location of the mean has an impact on stability probability: (I) represents the usual choice, $\boldsymbol{\mu} = [0, 0]^\top$; however other choices can clearly lead to (II) lower or (III) higher stability probabilities. B. The variances of a and b have an impact on stability probability: e.g. for fixed mean $\boldsymbol{\mu} = [0, 0]^\top$, taking smaller or larger variances leads to, respectively, (I) higher or (II) lower stability probabilities. C. The covariance between a and b has an impact on stability: e.g. for fixed mean $\boldsymbol{\mu} = [0, 0]^\top$, whether a and b covary negatively or positively leads to, respectively, (I) higher or (II) lower stability probabilities.

obtainable as Jacobians will necessarily be associated with non-zero density by such a defined h , therefore limiting its relevance.

An appropriate choice of h is thus vital. In particular, choosing h for the sake of mathematical convenience can only provide limited insight, if doing so comes at the cost of sacrificing realism. The so-called ‘diversity-stability debate’ [1] arose because general conclusions about stability were drawn from RMEs that were either specific to a particular system [17, 22, 23, 24, 25, 26, 27, 28] or chosen for mathematical convenience [11, 12] — e.g. to invoke the circular law [2, 7, 13] — yielding results unlikely to be meaningful for any interesting realistic system [29].

Here we show that the dichotomy resulting from the use of different RMEs can be overcome by constructing RMEs that are appropriate for, and conditioned on, the properties of Jacobian matrices of real systems. We also show that such RMEs should not be used to draw general conclusions regarding other systems than the ones they were built for.

2.3. Monte Carlo Estimates of Stability Probability

We consider autonomous ODE systems of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t); \boldsymbol{\theta}),$$

where $\mathbf{x}(t) = [x_1(t), \dots, x_p(t)]^\top$ is the vector of state variables at time t , $\boldsymbol{\theta}$ is the vector of parameters, and $\mathbf{f}(\mathbf{x}(t); \boldsymbol{\theta}) = [f_1(\mathbf{x}(t); \boldsymbol{\theta}), \dots, f_p(\mathbf{x}(t); \boldsymbol{\theta})]^\top$. By definition, an *equilibrium point*, \mathbf{x}_θ^* , of the system has the property:

$$\mathbf{f}(\mathbf{x}_\theta^*; \boldsymbol{\theta}) = \mathbf{0}.$$

We include the subscript $\boldsymbol{\theta}$ in our notation for the equilibrium point to emphasise that its location, existence, and stability will generally depend upon the particular values taken by the parameters. We denote by J_θ^* the Jacobian matrix of \mathbf{f} evaluated at \mathbf{x}_θ^* .

We can induce an ensemble, \mathcal{R} , of Jacobian matrices by specifying a distribution, \mathcal{F} , for the parameters $\boldsymbol{\theta}$: a collection of N parameter vectors $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$ sampled from \mathcal{F} defines an ensemble of corresponding matrices $J_{\theta^{(1)}}^*, \dots, J_{\theta^{(N)}}^*$. For any such RME, we may calculate a Monte Carlo estimate of the probability of stability, simply as the proportion of matrices that are stable; i.e. for which the leading eigenvalue has negative real part. That is, we obtain an estimate of the stability probability as,

$$\hat{P}(\text{stable}|\mathcal{R}) \approx \frac{1}{N} \sum_{i=1}^N \mathbb{I}(J_{\theta^{(i)}}^* \in \Lambda_S^n), \quad (3)$$

where $\mathbb{I}(X)$ is the indicator function, which is 1 if X is true and 0 otherwise.

This ensemble is designed to be the most realistic, since it fully takes into account the structure of the Jacobian matrix for the system. Hence it is only the choice of parameter distribution that determines the stability probability.

2.4. Conditional Random Matrix Ensembles

The previously defined stability probability is the probability of a system being stable, *conditional* on a given system architecture; as discussed in Section 2 and illustrated in Section 2.1 such architectures do not arise without a concrete context. However, the conditions for which the circular law is believed to hold lack this connection to reality, at least for mesoscopic systems. To study the effects of this context, which is encapsulated by the statistical properties of, and dependencies among, the entries in the Jacobian matrices $J_{\theta^{(1)}}^*, \dots, J_{\theta^{(N)}}^*$, we consider two further random matrix distributions, constructed by permutation of the entries of our original RME. First, we form a new matrix ensemble, $K_{(1)}^*, \dots, K_{(N)}^*$, in which the dependency between entries is broken. For each $\ell \in \{1, \dots, N\}$ and $(i, j) \in \{1, \dots, p\} \times \{1, \dots, p\}$ we set $(K_{(\ell)}^*)_{ij} = (J_{\theta^{(q)}}^*)_{ij}$, with q drawn uniformly at random from $\{1, \dots, N\}$. In this way, the marginal distribution of the ij -entries across the ensemble of K^* matrices is the same as the marginal distribution of ij -entries across the ensemble of J_θ^* matrices. Maintaining the marginal distributions ensures that the dependency between entries is the only quantity that we are altering:

in particular, the location of zeros in the matrix and the magnitudes of interaction strengths are maintained. We construct a further RME, $L_{(1)}^*, \dots, L_{(N)}^*$, where for each $\ell \in \{1, \dots, N\}$ and $(i, j) \in \{1, \dots, p\} \times \{1, \dots, p\}$, we set $(L_{(\ell)}^*)_{ij} = (J_{\theta^{(q)}}^*)_{rs}$, with q drawn uniformly at random from $\{1, \dots, N\}$, and r and s (independently) drawn uniformly at random from $\{1, \dots, p\}$. Now, the location of zeros in the matrix is no longer fixed; although the probability of an entry being zero is the same for the L^* matrices as for the J_{θ}^* 's and K^* 's. Moreover, each entry of the L^* matrices is *i.i.d.*. We henceforth refer to the J_{θ}^* matrices as the *FCS* (fully conditioned system) ensemble (most structure); the K^* matrices as the *independent* ensemble (intermediate structure); and the L^* matrices as the *i.i.d.* ensemble (least structure). We illustrate the properties of these three RMEs, and the methods for their construction, in Fig. 2.

To further our investigation we defined four more RMEs (which are presented in more detail in the *Appendix*). The first one will be referred to as the *independent normal* ensemble. It is constructed as follows: For each (i, j) , we fit an independent normal distribution to the ij -entries of the sampled Jacobians, $J_{\theta^{(1)}}^*, \dots, J_{\theta^{(N)}}^*$. That is, for each (i, j) , we calculate the mean,

$$\mu_{(i,j)}^{ind} = \frac{1}{N} \sum_{q=1}^N (J_{\theta^{(q)}}^*)_{ij},$$

and standard deviation,

$$\sigma_{(i,j)}^{ind} = s.d. \left\{ (J_{\theta^{(q)}}^*)_{ij} \right\}_{q=1}^N.$$

We then construct the new RME, $M_{(1)}^{ind}, \dots, M_{(N)}^{ind}$, where for each $\ell \in \{1, \dots, N\}$ and $(i, j) \in \{1, \dots, p\} \times \{1, \dots, p\}$, we set $(M_{(\ell)}^{ind})_{ij}$ to be a sample drawn from the univariate normal distribution with mean $\mu_{(i,j)}^{ind}$ and standard deviation $\sigma_{(i,j)}^{ind}$. By construction, the mean and standard deviation of the ij -entries across the ensemble of M^{ind} matrices are the same as the mean and standard deviation of the ij -entries across the ensemble of J_{θ}^* matrices (the *FCS* ensemble) and across the ensemble of K^* matrices (the *independent* ensemble).

A further ensemble is given by the *independent Pearson* ensemble. As in the *independent normal* case defined above, this new RME is defined by fitting a distribution to the ij entries of the sampled Jacobians, except that rather than using a normal distribution and just capturing the mean and standard deviation, we also capture the skewness and kurtosis of the ij -entries of the J_{θ}^* matrices. That is, in addition to $\mu_{(i,j)}^{ind}$ and $\sigma_{(i,j)}^{ind}$ defined earlier, we also calculate skewness

$$\gamma_{(i,j)}^{ind} = \text{skewness} \left\{ (J_{\theta^{(q)}}^*)_{ij} \right\}_{q=1}^N.$$

and kurtosis,

$$\kappa_{(i,j)}^{ind} = \text{kurtosis} \left\{ (J_{\theta^{(q)}}^*)_{ij} \right\}_{q=1}^N.$$

We then construct an RME, $M_{(1)}^{pear}, \dots, M_{(N)}^{pear}$, where for each $\ell \in \{1, \dots, N\}$ and $(i, j) \in \{1, \dots, p\} \times \{1, \dots, p\}$, we set $(M_{(\ell)})_{ij}^{pear}$ to be a sample drawn from a univariate Pearson distribution with mean $\mu_{(i,j)}^{ind}$, standard deviation $\sigma_{(i,j)}^{ind}$, skewness $\gamma_{(i,j)}$, and kurtosis $\kappa_{(i,j)}$. This RME thus shares many of the properties of the marginal distributions of ij -entries across the ensemble of J_{θ}^* matrices, but does not capture the dependencies between them.

The third additional RME will be referred to as the *i.i.d. normal* ensemble. This time we will not fit a distribution to the ij -entries of the J_{θ}^* matrices, but instead we fit a normal distribution, using the same technic that we used for the *independent normal* ensemble defined above, to the ij -entries of the L^* matrices (i.e. those from the *i.i.d.* ensemble).

Finally, we construct an RME that attempts to capture some of the dependencies between the entries of the J_{θ}^* matrices. We define $c(M)$ to be the vector obtained by concatenating the columns of the matrix M (and further define c^{-1} be the inverse operation, so that, for example, $c^{-1}(c(M)) = M$). Applying $c(\cdot)$ to the matrices from our *FCS* RME, we obtain N vectors of length $p \times p$, namely: $c(J_{\theta(1)}^*), \dots, c(J_{\theta(N)}^*)$. To these, we fit (by maximum likelihood) a $(p \times p)$ -variate normal distribution. We then sample N vectors, v_1, \dots, v_N , of length $p \times p$ from this distribution, and form a new ensemble $M_{(1)}^{mvn}, \dots, M_{(N)}^{mvn}$ by setting $M_{(q)}^{mvn} = c^{-1}(v_q)$. We will call this new ensemble the *multivariate normal* ensemble.

These new ensembles allow us to control which aspect of the structure of the *FCS* gives it its stability properties. For instance comparing the *independent normal* ensemble, the *independent Pearson* ensemble and the *independent ensemble* we can show the impact of the different moments of the distribution. The *multivariate normal* and *FCS* ensembles can be used for the same purpose in the case where dependencies are considered. More detail about the different RMEs is provided in the *Appendix*.

3. Results

3.1. RME Choice Determines Stability Assessment

Table 1. Stability Probabilities

	Tyson	SEIR	N&B	Lorenz
FCS	0.32109	1	1	1
Multivariate Normal	0.1105	0.72188	0.43744	0.92382
Independent	0.11151	0.56483	0.57161	0.7225
Univariate Pearson	0	0.58252	0.93296	0.43542
Univariate Normal	0.116	0.53502	0.26504	0.48366
I.i.d Normal	0.0377	0.14904	0.12702	0.14602
I.i.d.	0.03209	0.1371	0.12173	0.10419

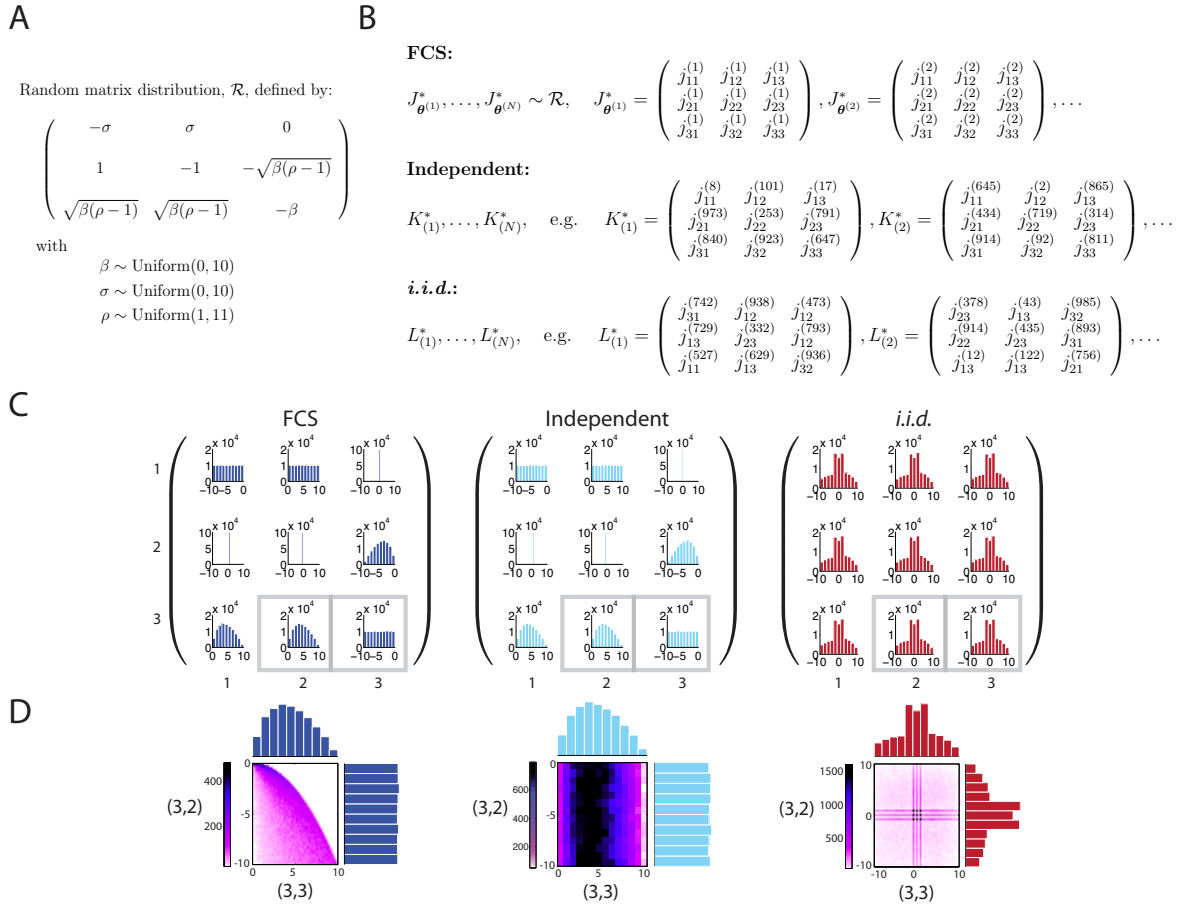


Figure 3. Random matrix ensembles (RMEs). A. For a given ODE system (e.g. the Lorenz equations) and equilibrium point, specifying a distribution for the parameters defines a random Jacobian matrix distribution. B. Samples from this distribution define the *FCS* matrix distribution; the *independent* and *i.i.d.* distributions are obtained from this by permuting elements as illustrated. $j_{kl}^{(m)}$ is the term in row k and column l obtained in the m^{th} sample from the random Jacobian matrix distribution \mathcal{R} . The marginal distributions for the elements of the matrices in the three distributions. In the *FCS* case, these reflect the parameter distributions and the expressions for the Jacobian entries presented in A; by construction, the marginals in the *independent* case are the same as for the *FCS*; while in the *i.i.d.* case, all entries have the same marginal distribution. D. We illustrate the joint distribution for two matrix entries: in the *FCS* case, the two entries exhibit dependency, whereas in the *independent* and *i.i.d.* cases, the joint is the product of the marginals.

Alternative: Stability, as stressed above and in the literature, is an issue in a wide variety of domains, and therefore we consider a set of systems that cover different qualitative behaviour of dynamical systems. The four ODE models that we consider have in common that the equilibrium points and Jacobians can be identified analytically, which makes analysis straightforward; they are: (i) the Lorenz system [30]; (ii) a model of the cell cycle [31]; (iii) a model of viral dynamics [32]; and (iv) an SEIR (susceptible-exposed-infective-recovered) population dynamics model [33]. In each case,

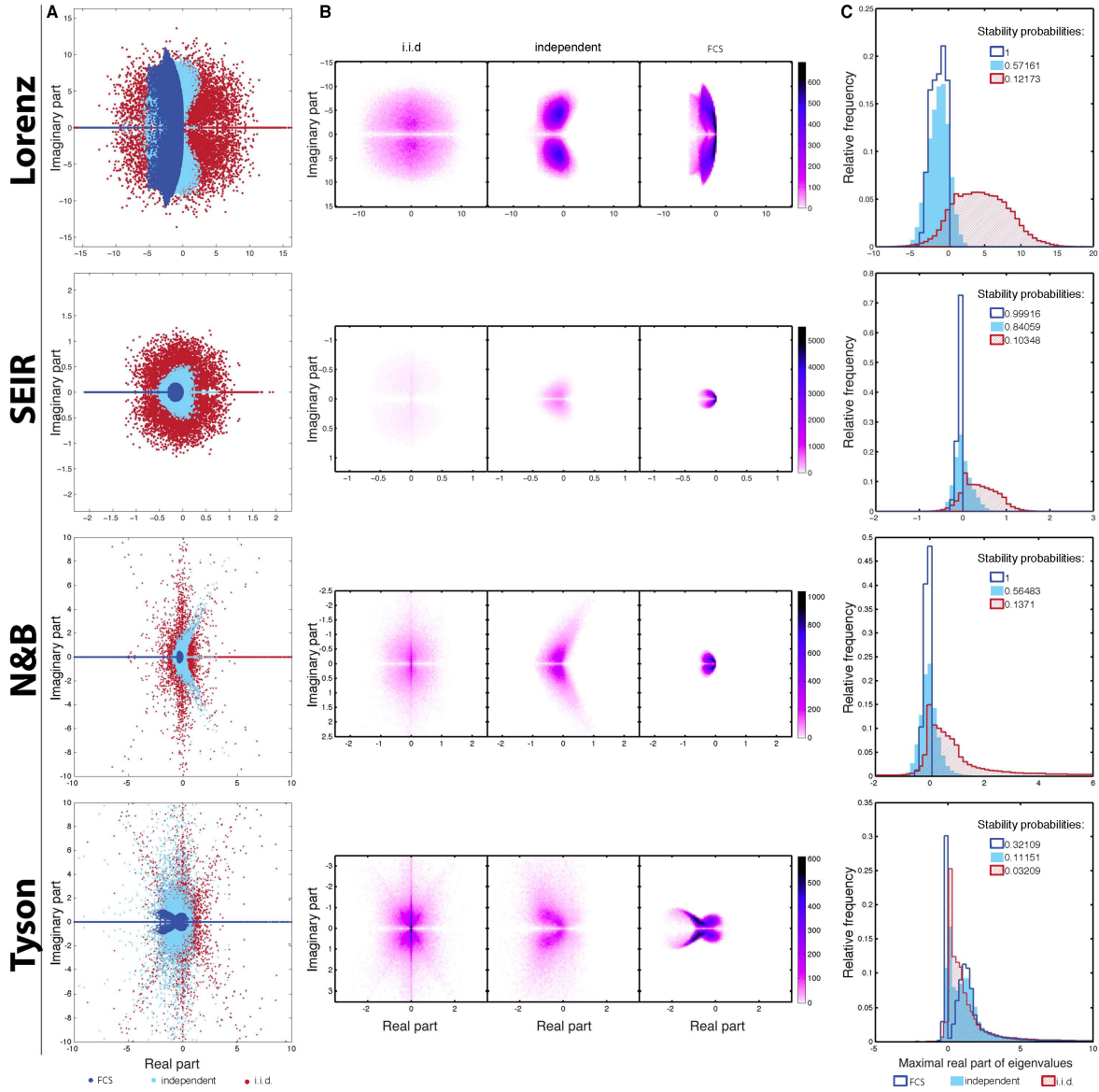


Figure 4. Stability results for the four example models. A. The eigenspectra for each model and random matrix distribution, shown as scatterplots. B. The eigenvalue distributions visualised using heat maps (to aid visualisation, we omit pure imaginary eigenvalues). C. The distributions of maximal eigenvalues together with the estimated stability probabilities.

we present results for physically or biologically feasible equilibrium points and generate 100,000 matrices from our RMEs in order to obtain Monte Carlo estimates of stability probabilities. Full details of these models and their corresponding RMEs are provided in the *Appendix*.

Fig. 4A shows the eigenspectra for the three RME regimes. While the *i.i.d.* eigenspectra are broadly circular, we observe diverse and decidedly non-circular shapes for the other two cases, highlighting the limitations of previous analyses based upon the circular law. Fig. 4B shows the density of eigenvalues in the complex plane for

the different models and RMEs. The eigenspectra distributions are typically much less dispersed for the *FCS* ensemble than for the other two. As shown in Fig. 4C this also leads to systematic differences in the real parts of the leading eigenvalues, which determine stability.

Table 1 shows that, whatever the model considered, the probability of stability of the *i.i.d.* ensemble is always very different from the probability stability of the *FCS*. The other RMEs, which all include more and more of the structure of the system in their construction, are getting closer to the *FCS*. In most cases the univariate Pearson ensemble had a probability stability closer to that of the *FCS* than the univariate normal, showing that considering more moments when building the RME improves the estimation of the stability of the system.

Monte Carlo estimates for the stability probabilities decrease as we decrease the amount of realism captured by the RME: failing to condition on the real-world heterogeneity and dependency present in the Jacobian can result in an unnecessarily pessimistic assessment of stability, even for these small systems. Considering RMEs in which we tightly control the mean, standard deviation, skewness and kurtosis of the marginal distributions of Jacobian entries, demonstrates that all of these properties also have an impact on stability.

3.2. Large Dynamical Systems Can be Stable

To illustrate the effects of inadequately capturing model structure and parameter dependencies on the stability probabilities of larger systems, we consider extensions to the SEIR model (Fig. 5A). We allow for multiple subpopulations of exposed (E) individuals (in the *Appendix* we also investigate extensions with heterogeneous infective subpopulations), enabling us to control system size. We again consider the three RMEs described above (see *Appendix* for full details). As we increase the size of the system, the probability of stability remains 1 in the *FCS* case, but rapidly diminishes in the *i.i.d.* case (Fig. 5B). The *independent* case is intermediate, indicating that not only can the dependence between matrix entries be important, but also their heterogeneity. Heterogeneity changes the location of the centre of the matrix p.d.f. h , and also how it stretches in different directions, which modifies the proportion of probability mass falling in the stable region, $\Lambda_S^n \subset M_n(\mathbb{R})$.

Fig. 5C shows how summaries of the distributions of leading eigenvalues change as we increase the number of exposed populations, with Fig. 5D providing corresponding density estimates for a selection of these numbers. In the *i.i.d.* case, the distributions and median values shift away from stable negative values toward unstable positive values. This is in stark contrast to the *FCS* case where, regardless of the number of exposed populations, the distribution of leading eigenvalues only has support on the negative real line (and hence we always have stability probability 1). Moreover, as we increase the number of exposed populations, the median value of the leading eigenvalue tends to become *more* negative. The *independent* case is again intermediate, with the median

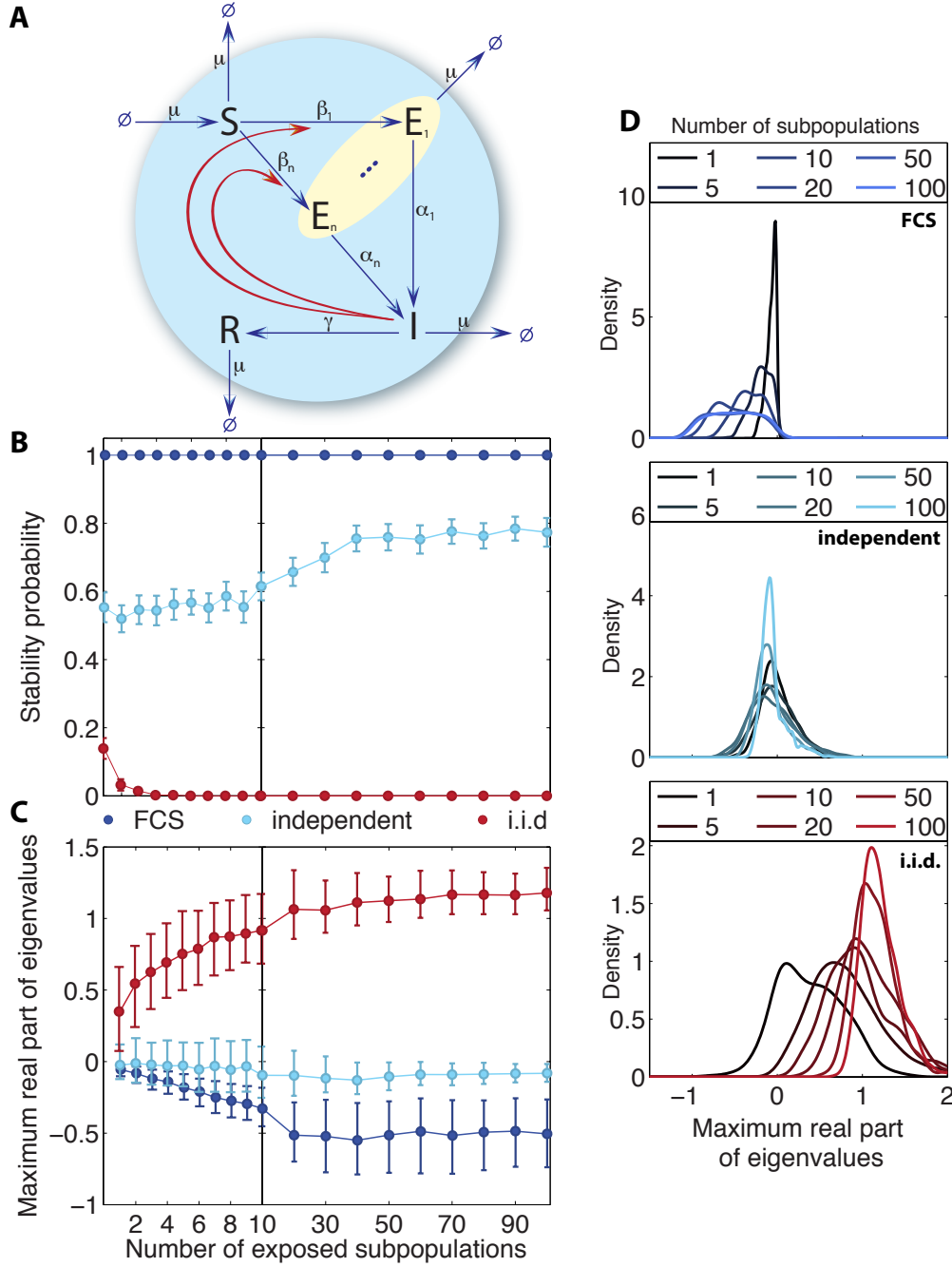


Figure 5. How stability changes with system size depends on the random matrix distribution. A. An extension of the SEIR model in which we model n exposed populations, E_1, \dots, E_n . B. Plot showing for each of the random matrix distributions how estimated stability probability changes as we increase the number of exposed populations. Bars denote ± 2 s.d. Monte Carlo error bars. C. Plot showing median (filled circle) and interquartile range (bars) for the distributions of leading eigenvalues. D. Density estimates for the distributions of leading eigenvalues.

value staying relatively constant.

Figures C3 and C4 in the Appendix, where we consider the effect of the *i.i.d. normal*, the *independent normal*, the *independent Pearson*, and the *multivariate normal* ensembles on these models, bring more evidence to our previous observations. As we would expect, the more of the underlying system’s structure that we capture using our chosen RME, the closer we get to the stability probability estimated using the *FCS* ensemble. The *i.i.d. normal* RME, which does not include any more structure than the *i.i.d.* ensemble, leads to similar stability probabilities to those obtained using that RME. The *independent normal*, which allows the heterogeneity of variances and means found in the real system to be described, yields stability probabilities closer to the *FCS* ensemble. The *independent Pearson*, which also takes into account information about the skewness and kurtosis of each entry, gets even closer to the *FCS* ensemble, and is very similar to the *independent* case. Finally, the *multivariate normal* RME (which allows us to model – in a simplistic fashion – some of the dependencies between the entries of the Jacobian) results in stability probabilities that are always closer to the *FCS* ensemble than those obtained using the *independent normal* RME. We found that the stability probabilities obtained using the *multivariate normal* RME are not consistently more different from *FCS* stability probabilities than those obtained using the *independent Pearson* RME. Thus, accounting for dependency between the Jacobian’s entries may, depending on the problem at hand, be more or less important than accounting for higher order properties of the marginal distributions of the entries.

4. Discussion

The stability of real-world systems — which is nearly ubiquitously observed — might seem perplexing in light of classical results in random matrix theory. By considering how random matrices can be made to reflect the properties of the Jacobian matrix of real dynamical systems it becomes possible to resolve and reconcile apparent contradictions in the literature.

In agreement with previous authors, our results demonstrate that stability probability can be affected by the mean and standard deviation of the entries in the Jacobian, as well as the dependencies between them [26, 27], and further show that properties such as the skewness and kurtosis of the entries can also have an impact. This is unsurprising: as is clear from Equation (2) and illustrated in Fig. 2, RMEs with different properties will result in different proportions of probability mass falling within the stable region, $\Lambda_S^n \subset M_n(\mathbb{R})$. Reported stability probabilities should therefore always explicitly acknowledge that they are conditioned on a particular choice of RME, which has to be carefully justified.

Whilst May’s mathematical study clearly shows that in some circumstances an increase in complexity can lead to instability[2], Haydon’s study highlights cases where complexity, in the shape of strong and numerous interconnections, is necessary to get higher stability[11]. Other examples where complex systems have been shown to be

stable can be found in the literature, in particular Kokkoris *et al.* show that different variances of the interaction strength will allow for different levels of complexity of the system whilst keeping the same stability[26]. However, none of these results can be generalised lightly. They in fact show that different systems are impacted differently by changes in complexity, and that no general prediction can be made.

Here, the *FCS* ensemble conditions on the model structure, so that the RME is defined through the distribution of model parameters. In our case, these distributions have been chosen by selecting plausible or interesting ranges for the parameters, and taking uniform distributions within these ranges (in the *Appendix* we consider alternative possibilities). In real applications, a natural choice for the distribution of model parameters would be provided in the Bayesian formalism by the posterior parameter distribution (or, if no data are available, by the prior). From this, we may obtain the posterior (or prior) predictive distributions [35] of the leading eigenvalues, from which we may derive the probability of stability. In this way, a truly realistic assessment of stability may be obtained for the system, in which we have conditioned on both our current understanding of the system architecture (encapsulated in our mathematical model) and our current uncertainty in the model's parameters.

Through our analyses, we have demonstrated that identifying any single property of the RME as being the general determinant of stability is misleading, except in some cases when the system has been very strictly defined [23, 12, 26, 27]. Stability probability is determined by the topology of the stable region, and how much probability mass is deposited within that region by the RME. This cannot be summarily deduced from any single property of the RME. At this stage it seems that system stability is system specific and that little can be gleaned from general approaches that will at best be uninformative if not entirely misleading. In particular, we cannot assess the probability of a system being stable based only on its size, diversity or complexity. It is especially important to keep this in mind as the stability of more complicated systems is considered (see e.g. [34, 36]).

This does not rule out the possibility that there are sets of rules or principles that could greatly shift the balance in favour of stability. Negative feedback, for example, is likely to lead to more stable behaviour (even for stochastic systems). It is in principle possible to condition on such local structures (in terms of correlated entries in the Jacobian) that may confer or contribute to overall stability of a system [37]. To apply such knowledge to real systems, however, would require a level of certainty about the underlying mechanisms that we currently lack for all but the most basic examples. But even in the presence of uncertainty about system structures, local negative feedback between species, for example, would tend to favour stability, whereas positive feedback (or merely the lack of negative feedback — a hallmark of stability in control theory [38]) would typically result in amplification of initially small perturbations to a system's behaviour.

Appendix A. Methods

Two main methods were used. The first used an analytical approach, whilst the second used a numerical approach. The first method was used on models 1 to 4, and on models 5 and 6 for $n = 1$ to 6. In these cases the number of samples used was 100000. The second method was used on model 5 and 6 for $n = 1..10$ and $n = 20, 30, \dots, 100$, this time, for computational reasons, the sample size was 1000.

Appendix A.1. The analytical method

For each of the models, the equilibrium points were found using Matlab's analytic equation solver, *solve*. The Jacobian was also described analytically using Matlab's *jacobian* function.

The different parameters were sampled from a uniform distribution using Matlab's *rand* function. The choice of the range of the parameters will be described in detail below.

Appendix A.1.1. Algorithm For each of the model, the equilibrium points and their stability for a different range of parameters are evaluated in the following way:

- (i) Define the system of ODEs.
- (ii) Solve $\dot{x}_i = 0, \forall i \in \{1, \dots, p\}$, where p is the number of species in the system and $x_i, i \in \{1, \dots, p\}$ is the set of species in our system.
- (iii) Compute the Jacobian matrix of our system.
- (iv) Sample a set of parameters for the system.
- (v) Evaluate the equilibrium points for that set of parameters.
- (vi) If the equilibrium points are biologically realistic, i.e. all of the species have a concentration that is positive or null, then evaluate the Jacobian matrix at those equilibrium points.
- (vii) Else, 'reject' that set of parameters and sample a new set.
- (viii) Reproduce steps 5 to 7 until the number of samples accepted reaches the number of samples wanted.
- (ix) For each Jacobian matrix obtained, compute the eigenvalues; consider their maximum real part, each system is considered as stable if and only if that maximum real part is strictly negative.
- (x) The probability of a system being stable is then the number of stable systems divided by the number of samples.

In order to evaluate the stability under the *independent* and *i.i.d.* conditions we add a step between steps 8 and 9: we process the Jacobian matrices by doing some permutations of the entries as described in section 2.4.

Appendix A.2. The numerical method

For each of the models, the equilibrium points were found using Matlab's equation solver, *solve*. The Jacobian was described analytically using Matlab's *jacobian* function. The different parameters were sampled from a uniform distribution using Matlab's *rand* function. The choice of the range of the parameters will be described in detail below.

This method was used only on the S(nE)IR and the SE(nI)R models for computational reason (the numerical method becoming too expensive with big n). This method works well in this case because for these models we know that there are 2 equilibrium points and they are easily identified, so we can easily segregate the cases corresponding to each of them when computing the probability of stability. The first equilibrium point, where the whole population is composed of recovered individuals, is not interesting because it is obviously stable, so we only consider the other equilibrium point in each system.

Appendix A.2.1. Algorithm For each of the model, the equilibrium points and their stability for a different range of parameters are evaluated in the following way:

- (i) Define the system of ODEs.
- (ii) Compute the Jacobian matrix of our system analytically.
- (iii) Sample a set of parameters.
- (iv) Solve $\dot{x}_i = 0, \forall i \in \{1, \dots, p\}$, where p is the number of species in the system and $x_i, i \in \{1, \dots, p\}$ is the set of species in our system, using the set of parameters sampled.
- (v) Only keep the equilibrium point which is not a population fully composed of recovered individuals and no other type of individuals.
- (vi) If the equilibrium points are biologically realistic, i.e. all of the species have a concentration that is positive or null, then evaluate the Jacobian matrix at those equilibrium points.
- (vii) Else, 'reject' that set of parameters and sample a new set.
- (viii) Reproduce steps 3 to 7 until the number of samples accepted reaches the number of samples wanted.
- (ix) For each Jacobian matrix obtained, compute the eigenvalues; consider their maximum real part, each system is considered as stable if and only if that maximum real part is strictly negative.
- (x) The probability of a system being stable is then the number of stable systems divided by the number of samples.

In order to evaluate the stability under the *independent* and *i.i.d.* conditions we add a step between steps 8 and 9: we process the Jacobian matrices by doing some permutations of the entries as described in section 2.4.

Appendix A.3. Parameter ranges

Two criteria were used to choose the range of the parameters: first they had to be realistic; second the range had to be small enough to allow a thorough sampling of the space obtained to be computationally tractable. In the cases when the ranges were fixed arbitrarily, we verified that choosing different ranges would not impact the qualitative results.

Appendix A.3.1. Model 1: The Lorenz system To get the results obtained in the main article, we sampled the parameters from the following ranges:

$$\beta \in [0, 10]$$

$$\rho \in [1, 11]$$

$$\sigma \in [0, 10]$$

ρ is considered to be bigger or equal to 1 in order to ensure more than just the origin as a equilibrium point, which makes our study more interesting.

Appendix A.3.2. Model 2: A model of the cell division cycle All the parameters were taken to be uniformly distributed between 0 and 1.

Appendix A.3.3. Model 3: The Nowak and Bangham model All the parameters were taken to be uniformly distributed between 0 and 1.

Appendix A.3.4. Models 4, 5 and 6: The SEIR and extended SEIR models

Model 4: the SEIR model All the parameters were taken to be uniformly distributed between 0 and 1.

Model 5: the $S(nE)IR$ model All the parameters were taken to be uniformly distributed between 0 and 1.

Model 6: the $SE(nI)R$ model All the parameters were taken to be uniformly distributed between 0 and 1.

- [1] K. S. McCann (2000) The diversity-stability debate. *Nature* 405(6783):228–233.
- [2] R. M. May (1972) Will a large complex system be stable? *Nature* 238(5364):413–414.
- [3] S. H. Strogatz (2001) *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering* (Studies in Nonlinearity).
- [4] C. Elton (1958) *The ecology of invasions by animals and plants* (Chapman & Hall).
- [5] R. MacArthur (1955) Fluctuations of Animal Populations and a Measure of Community Stability. *Ecology* 36(3):533.
- [6] M. R. Gardner, W. R. Ashby (1970) Connectance of Large Dynamic (Cybernetic) Systems: Critical Values for Stability. *Nature* 228(5273):784–784.
- [7] R. May (1973) *Stability and complexity in model ecosystems*. (Princeton University Press).

- [8] A. Edelman, N. R. Rao (1999) Random matrix theory. *Acta Numerica* 14:233–297.
- [9] M. L. Mehta (2004) *Random matrices* (Academic Press).
- [10] T. Tao, V. Vu, M. Krishnapur (2010) Random matrices: Universality of esds and the circular law. *The Annals of Probability* 38(5):2023–2065.
- [11] D. T. Haydon (2000) Maximally stable model ecosystems can be highly connected. *Ecology* 81(9):2631–2636.
- [12] A. M. Neutel (2002) Stability in Real Food Webs: Weak Links in Long Loops. *Science* 296(5570):1120–1123.
- [13] S. Allesina, S. Tang (2012) Stability criteria for complex ecosystems. *Nature* 483(7388):205–208.
- [14] L. R. Lawlor (1978) A Comment on Randomly Constructed Model Ecosystems. *The American naturalist* 112(984):445–447.
- [15] D. Tilman, J. A. Downing (1994) Biodiversity and stability in grasslands. *Nature* 367:363–365.
- [16] T. J. Givnish (1994) Does diversity beget stability? *Nature* 371(6493):113–114.
- [17] A. Roberts (1974) The stability of a feasible random ecosystem. *Nature* 251(5476):607–608.
- [18] T. A. Brody et al. (1981) Random-matrix physics: spectrum and strength fluctuations *Rev.Mod.Phys.* 53:385–477.
- [19] E. P. Wigner (1959) Statistical properties of real symmetric matrices with many dimensions (1959) in *Statistical Theories of Spectra* 188–198.
- [20] T. Tao (2012) *Topics in Random Matrix Theory*. American Mathematical Society.
- [21] M. R. Schroeder (2009) *Number Theory in Science and Communication, 5th edition*. Springer.
- [22] M. E. Gilpin (1975) Stability of feasible predator-prey systems. *Nature* 254(5496):137–139.
- [23] A. W. King, S. L. Pimm (1983) Complexity, Diversity, and Stability - a Reconciliation of Theoretical and Empirical Results. *The American naturalist* 122(2):229–239.
- [24] S. Pimm (1984) The complexity and stability of ecosystems. *Nature* 307:321–326.
- [25] K. McCann, A. Hastings, G. R. Huxel (1998) Weak trophic interactions and the balance of nature. *Nature* 395(6704):794–798.
- [26] G. D. Kokkoris, V. A. A. Jansen, M. Loreau, A. Y. Troumbis (2002) Variability in interaction strength and implications for biodiversity. *Journal of Animal Ecology* 71(2):362–371.
- [27] V. A. A. Jansen, G. D. Kokkoris (2003) Complexity and stability revisited. *Ecology Letters* 6(6):498–502.
- [28] M. Christianou, G. D. Kokkoris (2008) Complexity does not affect stability in feasible model communities. *Journal of Theoretical Biology* pp. 162–169.
- [29] M. R. Evans, V. Grimm, K. Johst, T. Knuuttila, R. de Langhe, et al. (2013) Do simple models lead to generality in ecology? *Trends in Ecology & Evolution* 28(10):578–583.
- [30] E. N. Lorenz (1963) Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences* 20(2):130–148.
- [31] J. J. Tyson (1991) Modeling the cell division cycle: cdc2 and cyclin interactions. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 88(16):7328–7332.
- [32] M. A. Nowak, C. Bangham (1996) Population dynamics of immune responses to persistent viruses. *Science (New York, NY)* 272(5258):74–79.
- [33] J. L. Aron, I. B. Schwartz (1984) Seasonality and period-doubling bifurcations in an epidemic model. *Journal of theoretical biology* 110(4):665–679.
- [34] A. G. Haldane, R. M. May (2011) Systemic risk in banking ecosystems. *Nature* 469(7330):351–355.
- [35] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC, second edition.
- [36] R. D. Wilkinson (2013) Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Stat.Appl.Genet.Mol.Biol.* 12(2):129–141.
- [37] T. W. Thorne, M. P. H. Stumpf (2007) Generating confidence intervals on biological networks. *BMC Bioinformatics* 8(1):467.
- [38] C. Cosentino, D. Bates (2011) *Feedback Control in Systems Biology*. (CRC Press).

Appendix A. Defining meaningful stability probabilities

In the previous section, we demonstrated the importance of accounting for the structure present in Jacobians derived from real ODE models when calculating stability probabilities. For similar reasons, it is also important to account for other properties such as the *feasibility* of equilibrium points (i.e. whether or not they are physically meaningful). Since arbitrary choices of the matrix p.d.f. h will lead to arbitrary stability probabilities (further illustrated in Fig. A1), it is vital that we instead consider *meaningful* choices for h that are conditioned on such properties. As in other studies (outlined in the main manuscript), here we do so by defining random matrix ensembles (RMEs) for specific models via distributions over model parameters.

Appendix A.1. Defining alternative RMEs

The estimated stability probability defined in the main paper is the probability of stability, *conditional* on a given system architecture (i.e. conditional on the structure and dependency in the Jacobian that arises from a particular model).

To study the consequences of neglecting or incompletely capturing this structure, we first consider two random matrix distributions constructed by permutation of the entries of our original RME. To allow us to probe further the effects of RME choice on estimated stability probabilities, we moreover consider some RMEs for which the marginal distributions of the Jacobian entries are constrained to have particular parametric forms. Finally, we consider an RME in which we make some attempt to capture the dependency between the entries of the Jacobian.

Appendix A.1.1. The independent ensemble First, we form a new matrix ensemble, $K_{(1)}^*, \dots, K_{(N)}^*$, in which the dependency between entries is broken. For each $\ell \in \{1, \dots, N\}$ and $(i, j) \in \{1, \dots, p\} \times \{1, \dots, p\}$ we set $(K_{(\ell)}^*)_{ij} = (J_{\theta^{(q)}}^*)_{ij}$, with q drawn uniformly at random from $\{1, \dots, N\}$. In this way, the marginal distribution of the ij -entries across the ensemble of K^* matrices is the same as the marginal distribution of ij -entries across the ensemble of J_{θ}^* matrices. Maintaining the marginal distributions ensures that the dependency between entries is the only quantity that we are altering: in particular, the location of zeros in the matrix and the magnitudes of interaction strengths are maintained.

Appendix A.1.2. The i.i.d. ensemble We construct a further RME, $L_{(1)}^*, \dots, L_{(N)}^*$, where for each $\ell \in \{1, \dots, N\}$ and $(i, j) \in \{1, \dots, p\} \times \{1, \dots, p\}$, we set $(L_{(\ell)}^*)_{ij} = (J_{\theta^{(q)}}^*)_{rs}$, with q drawn uniformly at random from $\{1, \dots, N\}$, and r and s (independently) drawn uniformly at random from $\{1, \dots, p\}$. Now, the location of zeros in the matrix is no longer fixed; although the probability of an entry being zero is the same for the L^* matrices as for the J_{θ}^* 's and K^* 's. Moreover, each entry of the L^* matrices is *i.i.d.*.

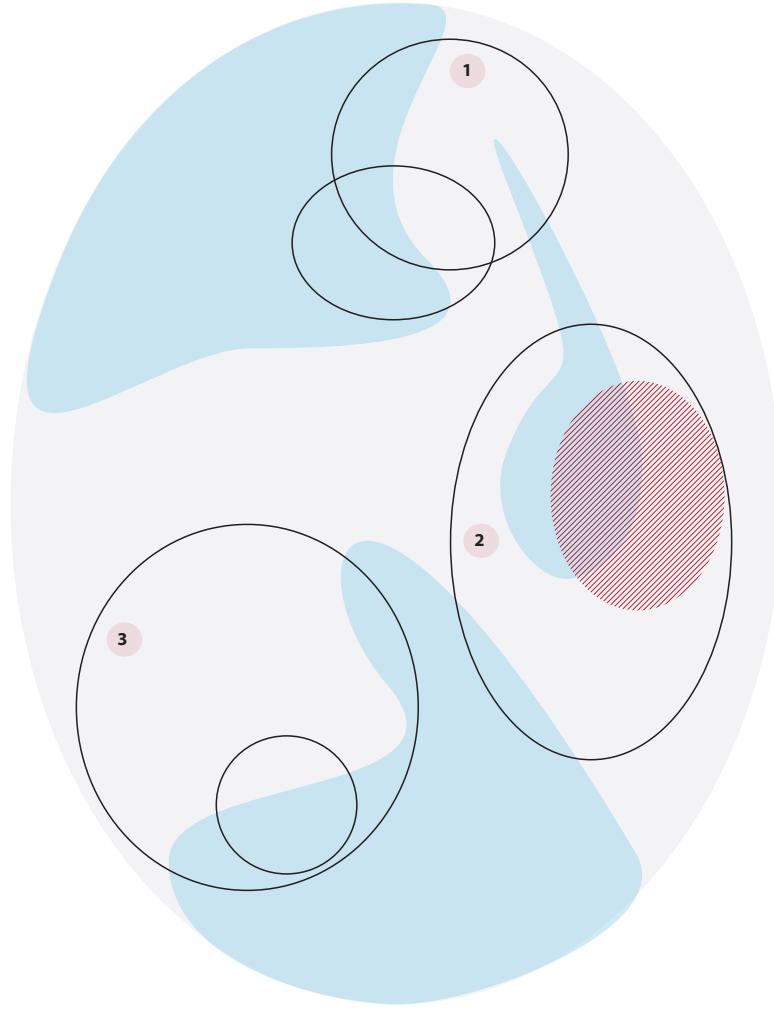


Figure A1. How the choice of RME can impact the probability of stability and usual mistakes made because of it. The grey ellipse represents $M_n(\mathbb{R})$, the blue areas are the stable areas of the system. Each black ellipse represents the space that could potentially be reached by sampling from an RME inferred from a specific ODE system. 1. The first error is to forget that one Jacobian can correspond to two (or more) different real dynamical systems, as illustrated by these two systems overlapping. 2. The second error is to forget that the biophysically feasible area, here represented with red stripes, can be different from the overall mathematically feasible area and have very different stability probability. 3. A third mistake is to think that any result obtained for one specific system with m links (or any characteristic) can be generalised to any system with m links. Here for instance the small circle represents the area of the space covered by a specific system with m links, the bigger circle on the other hand is the space covered by all the systems with m links. It is clear that probability of stability on each space is very different and thus generalising would be misleading.

Appendix A.1.3. The independent normal ensemble For each (i, j) , we fit an independent normal distribution to the ij -entries of the sampled Jacobians, $J_{\theta^{(1)}}, \dots, J_{\theta^{(N)}}$. That is, for each (i, j) , we calculate the mean,

$$\mu_{(i,j)}^{ind} = \frac{1}{N} \sum_{q=1}^N (J_{\theta^{(q)}}^*)_{ij},$$

and standard deviation,

$$\sigma_{(i,j)}^{ind} = s.d. \left\{ (J_{\theta^{(q)}}^*)_{ij} \right\}_{q=1}^N.$$

We then construct another RME, $M_{(1)}^{ind}, \dots, M_{(N)}^{ind}$, where for each $\ell \in \{1, \dots, N\}$ and $(i, j) \in \{1, \dots, p\} \times \{1, \dots, p\}$, we set $(M_{(\ell)}^{ind})_{ij}$ to be a sample drawn from the univariate normal distribution with mean $\mu_{(i,j)}^{ind}$ and standard deviation $\sigma_{(i,j)}^{ind}$. By construction, the mean and standard deviation of the ij -entries across the ensemble of M^{ind} matrices are the same as the mean and standard deviation of the ij -entries across the ensemble of J_{θ}^* matrices (the *FCS* ensemble) and across the ensemble of K^* matrices (the *independent* ensemble).

Appendix A.1.4. The independent Pearson ensemble As in the *independent normal* case (Appendix [Appendix A.1.3](#)), except that rather than just capturing the mean and standard deviation, we also capture the skewness and kurtosis of the ij -entries of the J_{θ}^* matrices. That is, in addition to $\mu_{(i,j)}^{ind}$ and $\sigma_{(i,j)}^{ind}$ defined earlier, we also calculate skewness

$$\gamma_{(i,j)}^{ind} = skewness \left\{ (J_{\theta^{(q)}}^*)_{ij} \right\}_{q=1}^N.$$

and kurtosis,

$$\kappa_{(i,j)}^{ind} = kurtosis \left\{ (J_{\theta^{(q)}}^*)_{ij} \right\}_{q=1}^N.$$

We then construct an RME, $M_{(1)}^{pear}, \dots, M_{(N)}^{pear}$, where for each $\ell \in \{1, \dots, N\}$ and $(i, j) \in \{1, \dots, p\} \times \{1, \dots, p\}$, we set $(M_{(\ell)}^{pear})_{ij}$ to be a sample drawn from a univariate Pearson distribution with mean $\mu_{(i,j)}^{ind}$, standard deviation $\sigma_{(i,j)}^{ind}$, skewness $\gamma_{(i,j)}$, and kurtosis $\kappa_{(i,j)}$. This RME thus shares many of the properties of the marginal distributions of ij -entries across the ensemble of J_{θ}^* matrices, but does not capture the dependencies between them.

Appendix A.1.5. The i.i.d. normal ensemble As in the *independent normal* case (Appendix [Appendix A.1.3](#)), except that rather than fitting to the ij -entries of the J_{θ}^* matrices, we instead fit to the ij -entries of the L^* matrices (i.e. those from the *i.i.d.* ensemble).

Appendix A.1.6. The multivariate normal RME Finally, we construct an RME that attempts to capture some of the dependencies between the entries of the J_{θ}^* matrices. We define $c(M)$ to be the vector obtained by concatenating the columns of the matrix M (and further define c^{-1} be the inverse operation, so that, for example, $c^{-1}(c(M)) = M$). Applying $c(\cdot)$ to the matrices from our *FCS* RME, we obtain N vectors of length $p \times p$, namely: $c(J_{\theta(1)}^*), \dots, c(J_{\theta(N)}^*)$. To these, we fit (by maximum likelihood) a $(p \times p)$ -variate normal distribution. We then sample N vectors, v_1, \dots, v_N , of length $p \times p$ from this distribution, and form a new ensemble $M_{(1)}^{mvn}, \dots, M_{(N)}^{mvn}$ by setting $M_{(q)}^{mvn} = c^{-1}(v_q)$.

Appendix B. Exemplar models

Appendix B.1. Model 1: The Lorenz system

We start by considering the Lorenz system, merely because it is simple and widely known.

We define it in the same way as it is in Lorenz's paper:

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= x(r - z) - y \\ \dot{z} &= xy - bz.\end{aligned}$$

We consider values of parameters for which the following equilibrium point exists:

$$[x, y, z]^{\top} = \left[\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1 \right]^{\top},$$

and consider the probability of stability for this point.

Appendix B.2. Model 2: A model of the cell division cycle

We use the model as defined in the phase plane analysis of the Tyson paper:

$$\begin{aligned}\dot{u} &= k_4(w - u)\left(\frac{k'_4}{k_4} + u^2\right) - k_6u \\ \dot{v} &= (k_1[aa]/[CT]) - k_2(v - w) - k_6u \\ \dot{w} &= k_3[CT](1 - w)(v - w) - k_6u \\ \dot{y} &= (k_1[aa]/[CT]) - k_2(v - w) - k_7(y - v)\end{aligned}$$

This system has only one fixed point, details of which can be found in the Matlab code (available upon request from the authors). We assess the stability probability for this point.

Appendix B.3. Model 3: The Nowak and Bangham model

As a second example to study, we consider a model of viral dynamics proposed by Nowak and Bangham (1996). This model describes the interactions between uninfected cells,

x , infected cells, y , and free virus particles, v :

$$\begin{aligned}\dot{x} &= \lambda - dx - \beta xv \\ \dot{y} &= \beta xv - ay \\ \dot{v} &= ky - uv.\end{aligned}$$

This system has two fixed points. We assess the stability probability for the more interesting of these, namely,

$$[x, y, v]^\top = \left[\frac{au}{\beta k}, \frac{\lambda\beta k - dau}{\beta ka}, \frac{\lambda\beta k - dau}{\beta ku} \right].$$

Appendix B.4. Models 4, 5 and 6: The SEIR and extended SEIR models

Appendix B.4.1. Presentation and background We consider two different extended versions of the SEIR model in which we allow either the Exposed population or the Infective population to have a collection of subpopulations.

Recall the standard SEIR model:

$$\begin{aligned}\dot{S} &= \mu - \beta SI - \mu S \\ \dot{E} &= \beta SI - (\mu + \alpha)E \\ \dot{I} &= \alpha E - (\mu + \gamma)I\end{aligned}$$

Here, S is the proportion of the population that is “Susceptible”, E is the proportion of the population that is “Exposed” (infected, but not yet infective), and I is the proportion of the population that is “Infective”. We omit (explicitly) modelling the proportion of the population that is “Recovered”, making use of the fact that we must have

$$\text{Susceptible} + \text{Exposed} + \text{Infective} + \text{Recovered} = 1.$$

The parameters of the system are:

- μ : the birth rate, which we assume is equal to the death rate;
- $1/\gamma$: the mean infective period;
- β : the contact rate;
- $1/\alpha$: the mean latent period of the disease.

This system has two fixed points. The first one corresponds to the extinction of the infection, i.e. the whole population is in the *recovered* state. The second fixed point is more interesting because the infection survives, its details can be found in the Matlab code (available upon request from the authors). We assess the stability probability for this second, more interesting, fixed point.

Appendix B.4.2. First extended model: the $S(nE)IR$ model We now introduce n subpopulations of the “Exposed” population, representing (for example) different age groups. We suppose that the mean latent period of the disease varies between these subpopulations. This model will henceforth be named ‘ $S(nE)IR$ ’.

This leads to the following model:

$$\begin{aligned}\dot{S} &= \mu - \sum_{j=1}^n \beta_j SI - \mu S \\ \dot{E}_i &= \beta_i SI - (\mu + \alpha_i) E_i, \quad \text{for } i = 1, \dots, n. \\ \dot{I} &= \sum_{j=1}^n \alpha_j E_j - (\mu + \gamma) I\end{aligned}$$

Here, $1/\alpha_i$ is the mean latent period of the disease in the i -th Exposed subpopulation, and $\beta = \sum_{j=1}^n \beta_j$ is the overall contact rate between “Susceptible” and “Infective” individuals. Note that for $n = 1$, we recover the standard SEIR model.

As in the SEIR model, this system has two fixed points. The first one corresponds to the extinction of the infection, i.e. the whole population is in the *recovered* state. The second fixed point is more interesting because the infection survives, its details can be found in the Matlab code (available upon request from the authors). We assess the stability probability for this second, more interesting, fixed point.

Appendix B.4.3. Second extended model: the $SE(nI)R$ model In the other model we introduce n subpopulations of the “Infective” population. We suppose that the mean infective period of the disease varies between these subpopulations. This model will henceforth be named ‘ $SE(nI)R$ ’.

This leads to the following model:

$$\begin{aligned}\dot{S} &= \mu - \sum_{j=1}^n \beta SI_j - \mu S \\ \dot{E} &= \sum_{j=1}^n \beta SI_j - (\mu + \sum_{j=1}^n \alpha_j) E \\ \dot{I}_i &= \alpha_i E - (\mu + \gamma_i) I_i, \quad \text{for } i = 1, \dots, n.\end{aligned}$$

Here, $1/c_i$ is the mean infective period of the disease in the i -th Infective subpopulation, and $a = \sum_{j=1}^n \alpha_j$ is the overall latent period of the disease. Here again, for $n = 1$, we recover the standard SEIR model.

As in the SEIR model, this system has two fixed points. The first one corresponds to the extinction of the infection, i.e. the whole population is in the *recovered* state. The second fixed point is more interesting because the infection survives, its details can be found in the Matlab code (available upon request from the authors). We assess the stability probability for this second, more interesting, fixed point.

Appendix C. Additional Results

Appendix C.1. Stability probabilities of the $SE(nI)R$ and $S(nE)IR$ models

Obtained using the analytical method with 100000 samples. Results are shown in Tables [C1](#) and [C2](#).

Table C1. Stability Probabilities of SnEIR models

	SEIR	S₂EIR	S₃EIR	S₄EIR	S₅EIR	S₆EIR
FCS	1	1	1	1	1	1
Independent	0.56483	0.54848	0.52312	0.50472	0.49197	0.47679
i.i.d.	0.1371	0.04142	0.01193	0.00356	0.00075	0.00014

Table C2. Stability Probabilities of SEnIR models

	SEIR	SE₂IR	SE₃IR	SE₄IR	SE₅IR	SE₆IR
FCS	1	1	1	1	1	1
Independent	0.56483	0.64252	0.649	0.65373	0.65538	0.65933
i.i.d.	0.1371	0.04791	0.01554	0.00483	0.00105	0.00025

These results are further illustrated in Fig. [C1](#).

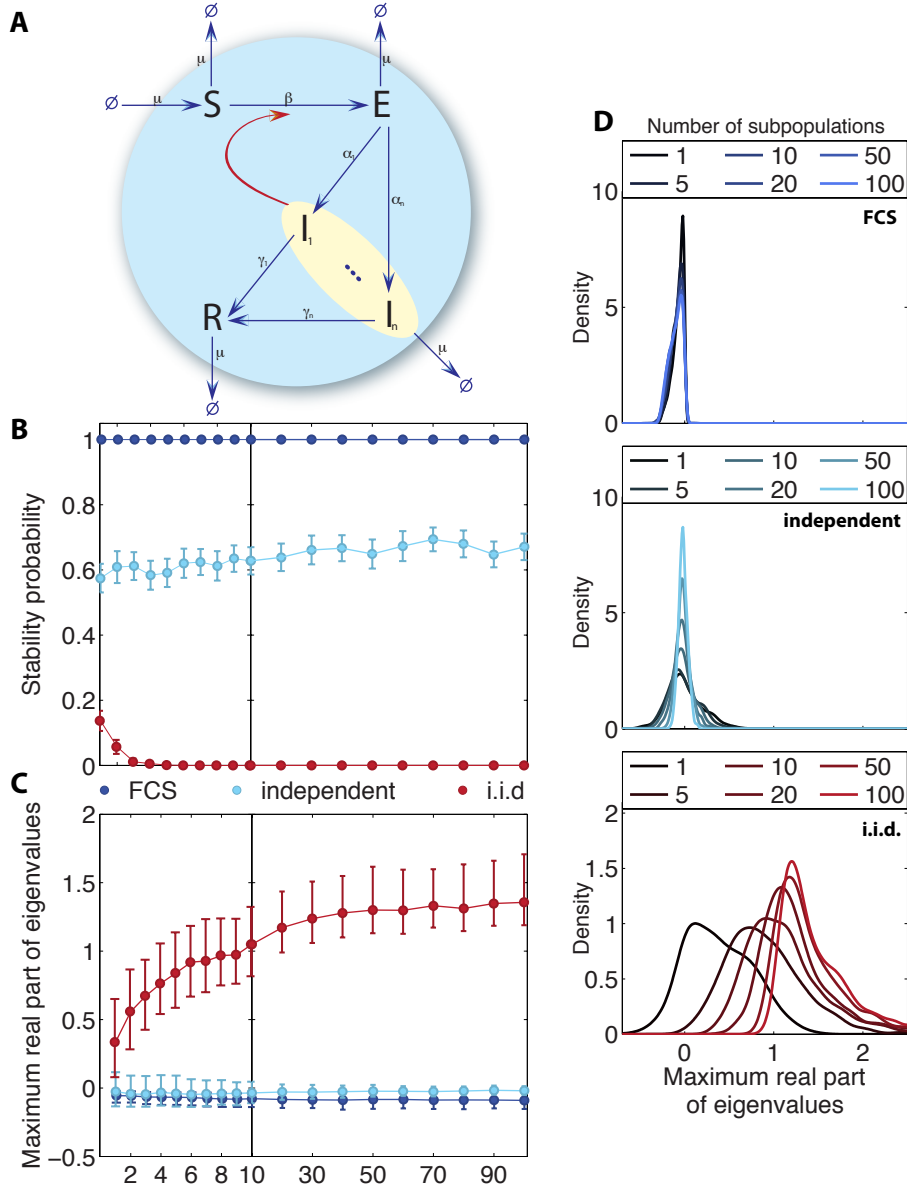


Figure C1. How stability changes with system size depends on the random matrix distribution. A. An extension of the SEIR model in which we model n infective populations, I_1, \dots, I_n . B. Plot showing for each of the random matrix distributions how estimated stability probability changes as we increase the number of exposed populations. Bars denote ± 2 s.d. Monte Carlo error bars. C. Plot showing median (filled circle) and interquartile range (bars) for the distributions of leading eigenvalues. D. Density estimates for the distributions of leading eigenvalues.

Appendix C.2. Additional ranges

In order to ensure that our results are not dependent on the range of the parameters considered we have considered different ranges for some of the models.

Appendix C.2.1. In the Lorenz model We will call Lorenz_i the Lorenz model with parameters sampled from uniform distributions with the following ranges:

$$\begin{aligned}\beta &\in [0, i] \\ \rho &\in [1, i + 1] \\ \sigma &\in [0, i]\end{aligned}$$

We took $i = 1, 10, 100, 1000, 10000$ and computed the probability of stability using the analytical method with 100000 samples. Results are shown in Table C3.

Table C3. Stability Probabilities of Lorenz model for different ranges

	Lorenz₁	Lorenz₁₀	Lorenz₁₀₀	Lorenz₁₀₀₀	Lorenz₁₀₀₀₀
FCS	1	0.99916	0.96825	0.96322	0.96271
Independent	0.7225	0.84059	0.81014	0.80335	0.80262
i.i.d.	0.10419	0.10348	0.10427	0.10426	0.10436

These results are further illustrated in Fig. C2.

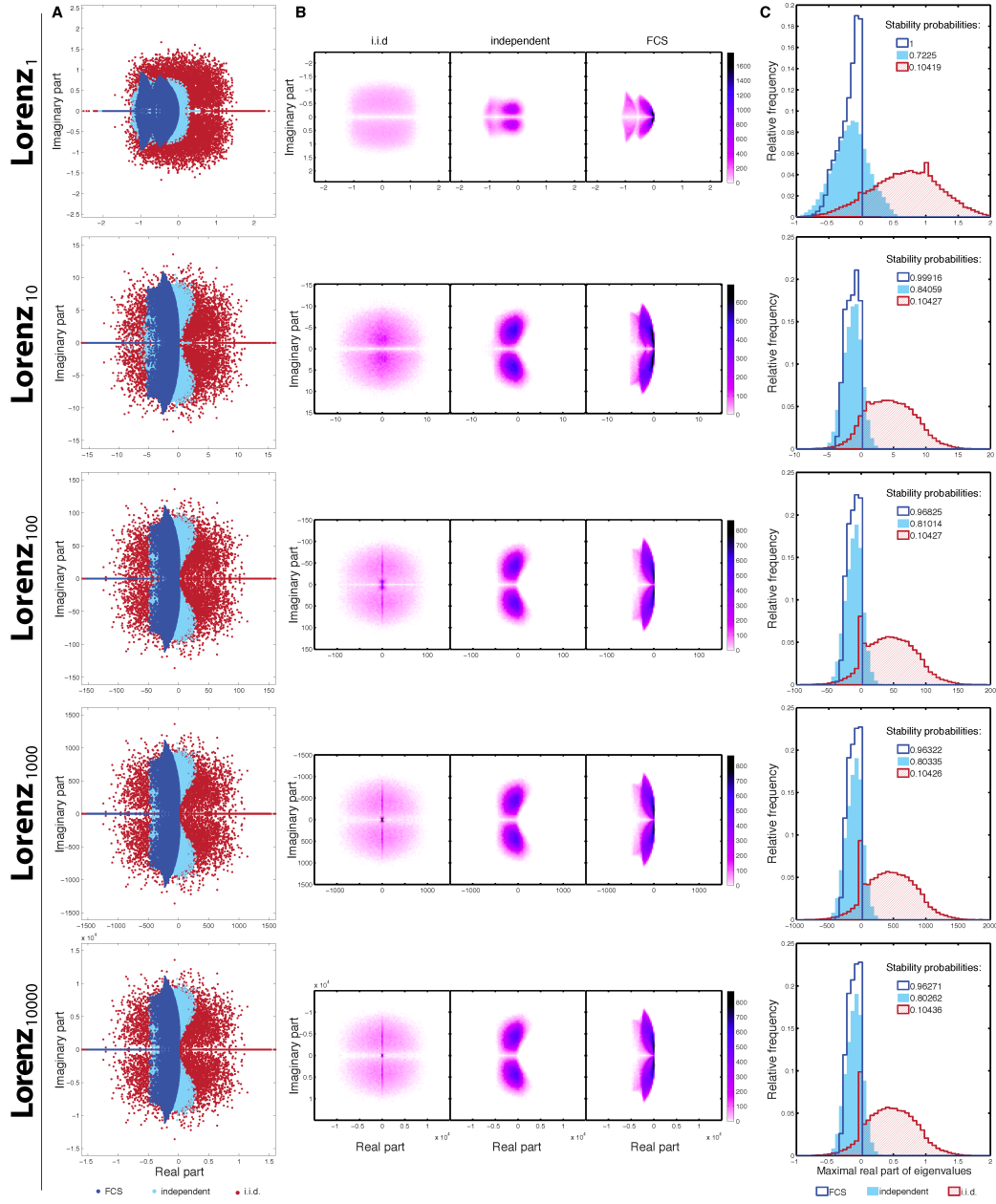


Figure C2. Stability results for additional ranges for the Lorenz model. A. The eigenspectra for each range and random matrix distribution, shown as scatterplots. B. The eigenvalue distributions visualised using heat maps (to aid visualisation, we omit pure imaginary eigenvalues). C. The distributions of maximal eigenvalues together with the estimated stability probabilities.

Appendix C.2.2. In the $S(nE)IR$ and $SE(nI)R$ systems We will call $S(nE)IR_i$ and $SE(nI)R_i$ the $S(nE)IR$ and $SE(nI)R$ models with parameters sampled from uniform distributions with the following ranges:

$$\mu \in [0, 1]$$

$$\gamma \in [0, i]$$

$$\beta \in [0, 1]$$

$$\alpha \in [0, i]$$

We took $i = 1, 10, 100$ and computed the probability of stability using the analytical method with 100000 samples. Results are shown in Tables C4–C8.

Table C4. Stability Probabilities of S_2EIR model for different ranges

	SE_2IR_1	SE_2IR_{10}	SE_2IR_{100}
FCS	1	1	1
Independent	0.64252	0.68502	0.69726
i.i.d.	0.04791	0.0496	0.04897

Table C5. Stability Probabilities of S_3EIR model for different ranges

	SE_3IR_1	SE_3IR_{10}	SE_3IR_{100}
FCS	1	1	1
Independent	0.649	0.70561	0.7183
i.i.d.	0.01554	0.01611	0.01701

Table C6. Stability Probabilities of S_4EIR model for different ranges

	SE_4IR_1	SE_4IR_{10}	SE_4IR_{100}
FCS	1	1	1
Independent	0.65373	0.71414	0.72522
i.i.d.	0.00483	0.0049	0.00478

Table C7. Stability Probabilities of S_5EIR model for different ranges

	SE_5IR_1	SE_5IR_{10}	SE_5IR_{100}
FCS	1	1	1
Independent	0.65538	0.71879	0.73039
i.i.d.	0.00105	0.00144	0.00122

Table C8. Stability Probabilities of S_6EIR model for different ranges

	SE_6IR_1	SE_6IR_{10}	SE_6IR_{100}
FCS	1	1	1
Independent	0.65933	0.72254	0.73651
i.i.d.	0.00025	0.00047	0.00045

Appendix C.3. Numerical $S(nE)IR$ and $SE(nI)R$ systems

The stability probabilities were estimated using 1000 samples, the error bars provided indicate plus/minus one standard deviation. Results are shown in Tables [C9](#) and [C10](#).

Table C9. Stability Probabilities of numerical SEnIR models

	numSE₁IR	numSE₂IR	numSE₃IR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.574 ± 0.023	0.606 ± 0.02	0.613 ± 0.024
i.i.d.	0.136 ± 0.015	0.0559 ± 0.0099	0.0118 ± 0.0051
	numSE₄IR	numSE₅IR	numSE₆IR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.584 ± 0.023	0.596 ± 0.023	0.628 ± 0.021
i.i.d.	0.0048 ± 0.0026	0.00092 ± 0.0013	0 ± 0
	numSE₇IR	numSE₈IR	numSE₉IR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.623 ± 0.021	0.614 ± 0.02	0.633 ± 0.019
i.i.d.	0 ± 0	0 ± 0	0 ± 0
	numSE₁₀IR	numSE₂₀IR	numSE₃₀IR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.626 ± 0.02	0.638 ± 0.022	0.662 ± 0.024
i.i.d.	0 ± 0	0 ± 0	0 ± 0
	numSE₄₀IR	numSE₅₀IR	numSE₆₀IR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.66 ± 0.02	0.651 ± 0.019	0.676 ± 0.019
i.i.d.	0 ± 0	0 ± 0	0 ± 0
	numSE₇₀IR	numSE₈₀IR	numSE₉₀IR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.688 ± 0.022	0.678 ± 0.02	0.642 ± 0.021
i.i.d.	0 ± 0	0 ± 0	0 ± 0
	numSE₁₀₀IR		
FCS	1 ± 0		
Independent	0.675 ± 0.022		
i.i.d.	0 ± 0		

Table C10. Stability Probabilities of numerical SnEIR models

	numS₁EIR	numS₂EIR	numS₃EIR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.551 ± 0.026	0.516 ± 0.021	0.549 ± 0.02
i.i.d.	0.139 ± 0.015	0.0315 ± 0.0089	0.0135 ± 0.0049
	numS₄EIR	numS₅EIR	numS₆EIR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.543 ± 0.02	0.559 ± 0.024	0.57 ± 0.02
i.i.d.	0.00192 ± 0.0018	0.00216 ± 0.002	0 ± 0
	numS₇EIR	numS₈EIR	numS₉EIR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.553 ± 0.022	0.588 ± 0.023	0.552 ± 0.023
i.i.d.	0 ± 0	0 ± 0	0 ± 0
	numS₁₀EIR	numS₂₀EIR	numS₃₀EIR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.618 ± 0.023	0.662 ± 0.024	0.7 ± 0.022
i.i.d.	0 ± 0	0 ± 0	0 ± 0
	numS₄₀EIR	numS₅₀EIR	numS₆₀EIR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.754 ± 0.018	0.755 ± 0.019	0.75 ± 0.018
i.i.d.	0 ± 0	0 ± 0	0 ± 0
	numS₇₀EIR	numS₈₀EIR	numS₉₀EIR
FCS	1 ± 0	1 ± 0	1 ± 0
Independent	0.778 ± 0.021	0.756 ± 0.021	0.783 ± 0.02
i.i.d.	0 ± 0	0 ± 0	0 ± 0
	numS₁₀₀EIR		
FCS	1 ± 0		
Independent	0.771 ± 0.021		
i.i.d.	0 ± 0		

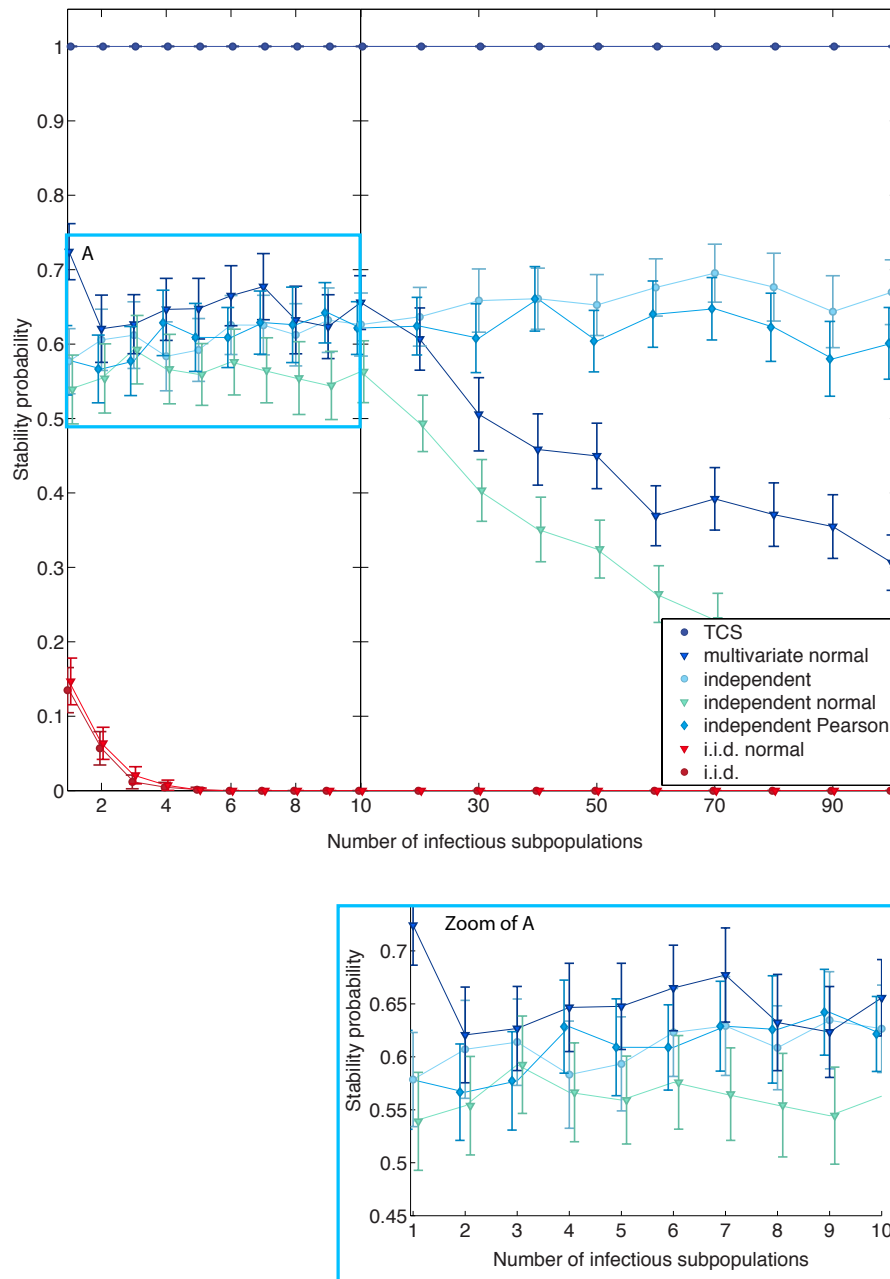


Figure C3. Stability probability of SE(nI)R models with different number of nodes, evaluated using different RMEs.

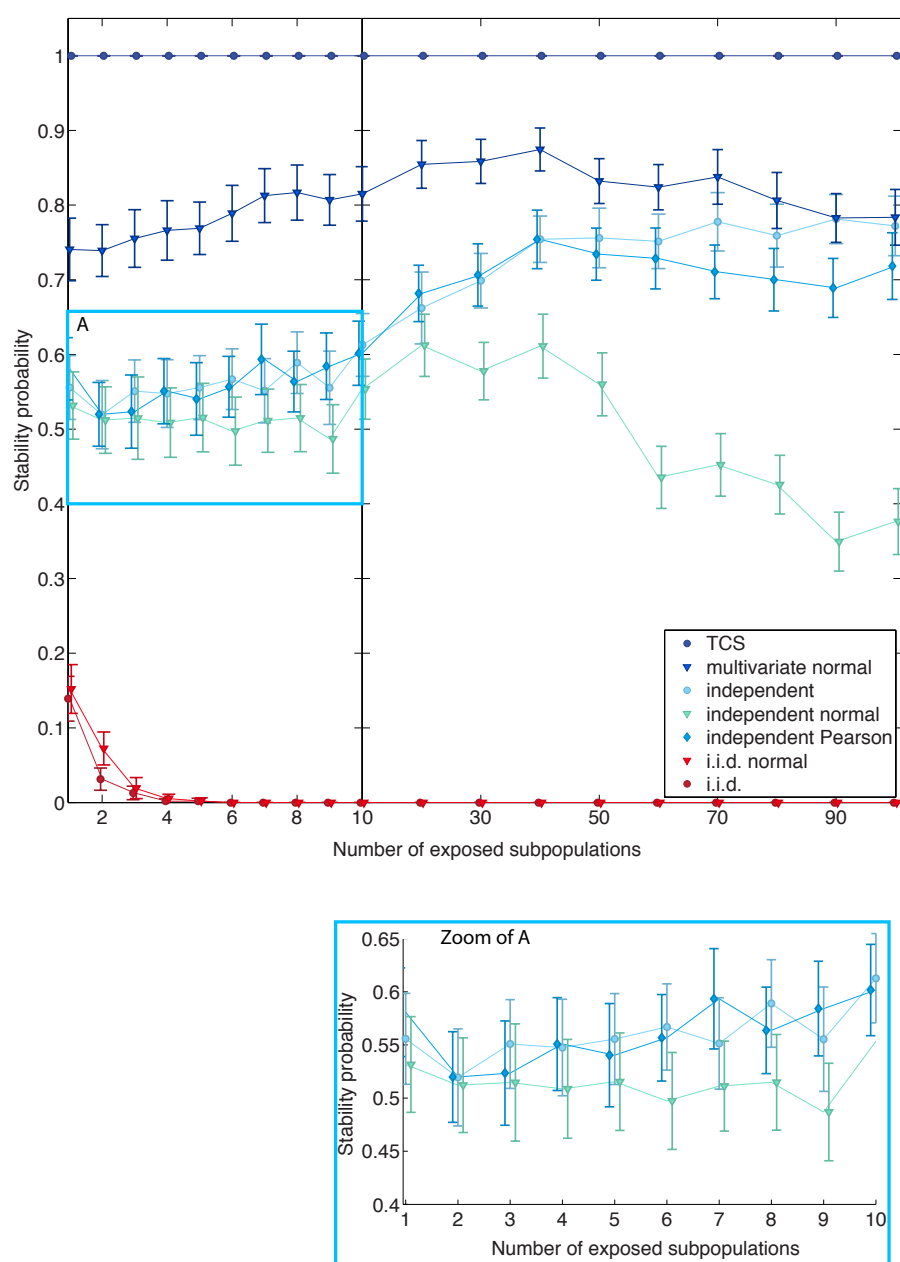


Figure C4. Stability probability of S(nE)IR models with different number of nodes, evaluated using different RMEs.